

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

Lipid analysis of CO<sub>2</sub>-rich subsurface aquifers suggests an autotrophy-based deep biosphere with lysolipids enriched in CPR bacteria.

### Permalink

<https://escholarship.org/uc/item/2gw9c9jc>

### Journal

The ISME journal, 14(6)

### ISSN

1751-7362

### Authors

Probst, Alexander J  
Elling, Felix J  
Castelle, Cindy J  
et al.

### Publication Date

2020-06-01

### DOI

10.1038/s41396-020-0624-4

Peer reviewed



# Lipid analysis of CO<sub>2</sub>-rich subsurface aquifers suggests an autotrophy-based deep biosphere with lysolipids enriched in CPR bacteria

Alexander J. Probst<sup>1,2</sup> · Felix J. Elling<sup>3,4</sup> · Cindy J. Castelle<sup>1,3</sup> · Qingzeng Zhu<sup>3</sup> · Marcus Elvert<sup>3</sup> · Giovanni Birarda<sup>5,6</sup> · Hoi-Ying N. Holman<sup>6</sup> · Katherine R. Lane<sup>1</sup> · Bethany Ladd<sup>7,9</sup> · M. Cathryn Ryan<sup>7</sup> · Tanja Woyke<sup>8</sup> · Kai-Uwe Hinrichs<sup>3</sup> · Jillian F. Banfield<sup>1</sup>

Received: 20 November 2018 / Revised: 5 February 2020 / Accepted: 25 February 2020 / Published online: 13 March 2020  
© The Author(s) 2020. This article is published with open access

## Abstract

Sediment-hosted CO<sub>2</sub>-rich aquifers deep below the Colorado Plateau (USA) contain a remarkable diversity of uncultivated microorganisms, including Candidate Phyla Radiation (CPR) bacteria that are putative symbionts unable to synthesize membrane lipids. The origin of organic carbon in these ecosystems is unknown and the source of CPR membrane lipids remains elusive. We collected cells from deep groundwater brought to the surface by eruptions of Crystal Geyser, sequenced the community, and analyzed the whole community lipidome over time. Characteristic stable carbon isotopic compositions of microbial lipids suggest that bacterial and archaeal CO<sub>2</sub> fixation ongoing in the deep subsurface provides organic carbon for the complex communities that reside there. Coupled lipidomic-metagenomic analysis indicates that CPR bacteria lack complete lipid biosynthesis pathways but still possess regular lipid membranes. These lipids may therefore originate from other community members, which also adapt to high in situ pressure by increasing fatty acid unsaturation. An unusually high abundance of lysolipids attributed to CPR bacteria may represent an adaptation to membrane curvature stress induced by their small cell sizes. Our findings provide new insights into the carbon cycle in the deep subsurface and suggest the redistribution of lipids into putative symbionts within this community.

## Introduction

The most prominent characteristic of the deep continental subsurface is the absence of sunlight. However, the diversity of subsurface ecosystems is manifold. Physicochemical characteristics, as well as the availability of electron donors

These authors contributed equally: Alexander J. Probst, Felix J. Elling

**Supplementary information** The online version of this article (<https://doi.org/10.1038/s41396-020-0624-4>) contains supplementary material, which is available to authorized users.

✉ Alexander J. Probst  
alexander.probst@uni-due.de

✉ Felix J. Elling  
felix\_elling@fas.harvard.edu

✉ Kai-Uwe Hinrichs  
khinrichs@uni-bremen.de

✉ Jillian F. Banfield  
jbanfield@berkeley.edu

<sup>1</sup> Department of Earth and Planetary Science, University of California, Berkeley, CA 94720, USA

<sup>2</sup> Institute for Environmental Microbiology and Biotechnology, Department of Chemistry, University of Duisburg-Essen, Essen, Germany

<sup>3</sup> MARUM Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany

<sup>4</sup> Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02138, USA

<sup>5</sup> Elettra-Sincrotrone Trieste, Strada Statale 14-km 163,5 Basovizza, 34149 Trieste, Italy

<sup>6</sup> Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

<sup>7</sup> Department of Geoscience, University of Calgary, Calgary, AB T2N 1N4, Canada

<sup>8</sup> DOE Joint Genome Institute, Walnut Creek, MA, USA

<sup>9</sup> Present address: Department of Earth, Ocean, and Atmospheric Sciences, University of British Columbia, Vancouver, Canada

and acceptors shape different microbial communities within these ecosystems (e.g., Refs. [1, 2]). In some environments, the availability of fossil organic matter, burial depth, and temperature exert strong control on community structure [3–5]. Other subsurface environments have low availability of buried organic matter. In such environments, genomic analyses suggest that in situ CO<sub>2</sub> fixation supports microbial communities [6–11]. Most subsurface environments may be sustained by fixed carbon from multiple sources, and the relative importance of in situ CO<sub>2</sub> fixation has been difficult to ascertain [12].

The candidate phyla radiation (CPR) of bacteria is a monophyletic group [13], which includes enigmatic small-celled microbes [14] that appear to be abundant predominantly in the subsurface [15]. Cocultures of CPR bacteria indicate that some are symbionts of other bacteria and heavily depend on their hosts for basic resources [16]. To date, none of the reconstructed CPR genomes encode for a complete fatty acid (FA)-based lipid biosynthesis pathway [15]. Other putative bacterial and archaeal symbionts from different branches of the tree of life also do not encode for their own lipid biosynthesis pathway [17–19] and at least one hyperthermophilic episymbiont (*Nanoarchaeum equitans*) has been suggested to acquire its lipids from the host archaeon [20]. However, the origin and types of lipids used by CPR bacteria remain elusive.

Analysis of the stable carbon isotopic ratios of lipid molecules has enabled researchers to track carbon flow through communities. For instance, it was shown that archaea growing in syntrophy with sulfate-reducing bacteria mediate the anaerobic oxidation of methane [21, 22]. This analysis was possible because the consortia were based on simple bacterial and archaeal assemblages that produce diagnostic lipid types. In another study, the stable carbon isotope ratios of methane and lipids were used to track the flow of carbon from methane into the two species thought to be present based on rRNA sequence profiling [23]. Coupled lipidomic, tag sequencing, and isotopic analyses also allow spatiotemporal tracking of carbon flow through complex microbial communities [24, 25]. However, the power of this approach is limited when microbial communities contain numerous organisms that produce unknown lipid molecules [26]. In fact, lack of information about the types of lipids produced by uncultivated organisms remains a major gap in microbial ecology.

A recent large-scale environmental genomics survey of subsurface microbial ecosystems within the Colorado Plateau, USA, provided evidence for a depth-based distribution of organisms affiliated with more than 100 different phylum-level lineages [12]. Samples were acquired from groundwater that erupted through the cold (i.e., non-thermal), CO<sub>2</sub>-driven Crystal Geyser. During the eruption cycle groundwater was sourced from different depths,

enabling the assignment of organisms to their respective depths. Genomic resolution of the tracked organisms linked three different carbon fixation pathways to groundwater from different depths. However, a major question remains regarding the extent to which autotrophic organisms provide organic carbon to these complex microbial communities. Further, the types and sources of lipids used to construct the cell envelope of CPR bacteria remain elusive. We postulated that clues regarding the types of lipids produced by uncultivated bacteria and archaea could be addressed by correlation-based analyses so long as sufficient numbers of samples were defined in terms of the abundances of the microorganisms present and overall lipid compositions of the same samples were available. Here, we use coupled metagenomic-lipidomic data sets to test this approach and to resolve the importance of autotrophy as the source of organic carbon in the studied environment.

## Material and methods

### Sampling scheme

Samples for lipid analyses were retrieved by collecting cells from groundwater sampled from the Crystal Geyser ecosystem onto a 0.1-μm teflon filter (Gravertech 10" MEM-TREX-HFE). Filters with biomass were immediately frozen on dry ice. One post-0.2-μm fraction was also collected to enrich for organisms of the CPR and DPANN radiations (sample ID 26, beginning of the recovery phase of the geyser). The samples span an entire cycle of the geyser, which lasted for ~5 days [12]. Collection for each metagenomic sample proceeded for around 4 h (141 L, SD 31%, Table S1). Collection of lipid samples proceeded simultaneously, but the collection time was around 8 h (114–338 L, Table S2) so there are half as many lipid samples as metagenome samples. The sampling scheme details are presented in Fig. S1. For infrared analysis coupled to metagenomics, one additional size-fractionated sample (first 0.2 μm, then 0.1 μm filtration) was included, which was collected during the recovery phase of the geyser in August 2014 and had been analyzed regarding its genomes earlier [12]. Details on samples and SRA accessions are provided in the Supplementary information.

### Sampling and isotopic analysis of dissolved inorganic carbon

Twenty-four groundwater samples were collected from about 8.5 m below ground surface in the geyser borehole using a peristaltic pump and copper pipe. Samples were collected in 12 mL glass vials. The vials were flushed with fresh geyser water and were filled underwater in a bucket

that was overflowing with groundwater to avoid atmospheric contact; this was confirmed by gas chromatography analyses that did not detect contamination by atmospheric gases (N<sub>2</sub>, O<sub>2</sub>, or Ar; unpublished data). The stable carbon isotopic composition of the dissolved inorganic carbon was analyzed by Continuous Flow Isotope Ratio Mass Spectrometry (CF-IRMS) using a Thermo Finnigan GasBench coupled to a DeltaV<sup>Plus</sup>. Water pressure, temperature, and electrical conductivity were measured in situ at the same depth using a Solinst LTC Levellogger Edge.

## Lipidomics

Methods for lipid extraction and analysis are described in detail in the Supplementary information (sample overview is given in Table S2). In brief, lipids were extracted using a modified Bligh and Dyer method [27] after addition of an internal standard. Archaeal and bacterial intact polar lipids (IPLs; for structures see Fig. S2) were quantified using a Dionex Ultimate 3000 ultra-high-performance liquid chromatography (UPLC) system connected to a Bruker maXis Ultra-High Resolution quadrupole time-of-flight mass spectrometer equipped with an electrospray ion source operating in positive mode (Bruker Daltonik, Bremen, Germany). Lipids were separated using normal phase UPLC on an Acquity UPLC BEH Amide column (1.7  $\mu$ m, 2.1  $\times$  150 mm; Waters Corporation, Eschborn, Germany) maintained at 40 °C as described in Ref. [28]. For isotopic analysis, IPLs were separated from free core lipids using semi-preparative high-performance liquid chromatography. For mass spectrometric analysis of previously uncharacterized IPLs (see Figs. S3 and S4). Ether cleavage and saponification were performed on the IPL fractions to release isoprenoid hydrocarbons and FA, respectively. The stable carbon isotopic compositions of these compounds were analyzed using gas chromatography–IRMS. Fourier-transform infrared (FTIR) spectromicroscopy was performed to detect lipids in intact cells. The FTIR system consisted of a Hyperion 3000 Infrared-Visible microscope coupled to a Vertex70V interferometer (Bruker Optics—Billerica, MA). For FTIR analysis, cells were deposited on a double-side-polished silicon slide and dried with a gentle nitrogen gas stream in a biological safety cabinet. Lipid identification was achieved by comparing spectra from samples and dry films of lipid standards.

## Metagenomics

Methods for DNA extraction and metagenomic sequencing are described in Ref. [12]. In brief, DNA was extracted from filters using the MoBio PowerMax Soil DNA isolation kit, and library preparation and sequencing were performed at the Joint Genome Institute (details on extracted DNA, type

of library and sequencing are provided in Ref. [12]). Quality filtered reads (<https://github.com/najoshi/sickle>, <https://sourceforge.net/projects/bbmap>) were assembled using IDBA\_UD [29], genes were predicted using prodigal (meta-mode; [30]). Coverage of scaffolds was calculated using bowtie2 (sensitive) [31]. Taxonomy of scaffolds was determined by searching proteins against an in-house database.

## Tracking taxa across time using ribosomal protein S3

In order to get a near-complete picture of specific taxa present in the samples, we extracted ribosomal protein S3 (rpS3) sequences from all assembled scaffolds >1 kb using separately designed HMMs for archaea, bacteria, and eukaryotes ([https://github.com/AJProbst/rpS3\\_trckr](https://github.com/AJProbst/rpS3_trckr)). The extracted amino acid sequences were clustered at 99% identity (collapsing most of the strains of the same species [32]) and the longest scaffold bearing a representative rpS3 sequence was obtained for each cluster. Using read mapping (bowtie2, [31]) and allowing a maximum of three mismatches per read (according to the 99% identity of the de-replicated rpS3 sequences), the relative abundance of each selected rpS3 scaffold was calculated across all samples. The breadth (i.e. how much of the sequence of a scaffold is covered) of the scaffolds was calculated in each sample. To call a rpS3 sequence present in a sample, it had to be either assembled or have a breadth of at least 95% of the entire scaffold in a sample. Since we worked with scaffolds, we did not consider ambiguous bases for calculating the breadth. The rpS3 sequences were taxonomically annotated against a combined database from previous publications [12, 33, 34], which was de-replicated at 99% rpS3 identity. Taxonomic assignments were performed with similarity cutoffs as described earlier:  $\geq 99\%$  for species,  $\geq 95\%$  for genus, and  $\geq 90\%$  for family level. Lower percentages were assigned to phylum or domain level (<50%).

## Statistical analysis to correlate taxa abundance with IPLs

Relative abundance measures of rpS3 genes were correlated (Pearson correlation) with relative abundance measure of IPLs if the rpS3 gene/the IPL species was present in at least 7 out of 14 samples. Resulting *p* values underwent false discovery correction using the Bonferroni procedure and these *q* values were then weighted by division of the *q* value with the percent relative abundance of the rpS3 gene. Each lipid was allowed to be assigned to only one organism (with the best score). This assignment of lipids to rpS3 genes considers that highly abundant organisms are more likely to

be detected in lipid analyses than low abundant organisms. IPL signatures were co-correlated (Bonferroni-corrected  $p$  value < 0.005) and lipid species that correlated with other lipids were identified for further analyses. These co-correlated lipid species, as well as the correlation of rpS3 genes and lipid species were used to construct a network (code is available under [https://github.com/AJProbst/lip\\_metgen](https://github.com/AJProbst/lip_metgen)) and visualized in Cytoscape. Primary lipids were assigned based on direct correlation of lipids with organisms, secondary lipids were assigned based on a correlation with primary lipids. Lipids were classified as unspecific if the secondary lipid correlated with two primary lipids of different organisms.

## Binning of genomes

rpS3 genes that were not found in existing genomes [12, 35] were identified based on a similarity (<98% [36]) and searched for in the respective metagenomes. Genomes containing these rpS3 sequences were binned using a consensus of guanine–cytosine content, coverage and taxonomy information in the ggKbase platform [37]. Genomes were subsequently curated with ra2 [13] for scaffolding errors. Genomes have been deposited at DDBJ/ENA/GenBank under the accessions SAMN13287258–462 (Umbrella BioProject PRJNA602879).

## Genomic analysis of lipid biosynthesis pathways in CPR genomes

Protein sequences were annotated from USEARCH (–ublast) searches against UniProt, UniRef100 [38], and KEGG databases [39] and uploaded to ggKbase (<https://ggkbase.berkeley.edu>). Based on existing annotations target proteins involved in bacterial FA, isoprenoids, and lipids biosynthesis were identified in CPR genomes and can be accessed using the following link: [https://ggkbase.berkeley.edu/genome\\_summaries/1491-Bacterial\\_membrane\\_lipids\\_AJP](https://ggkbase.berkeley.edu/genome_summaries/1491-Bacterial_membrane_lipids_AJP).

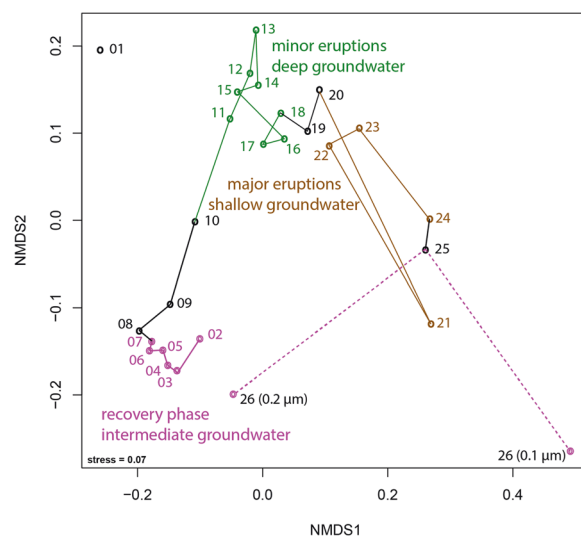
## Results and discussion

### Microbial community profile based on marker genes

We de novo assembled 27 metagenome samples, the reads from which were previously used in a study that involved mapping to 505 genomes reconstructed from prior data sets to link organisms to groundwater of different depths [12]. In the current study, we extracted assembled sequences of rpS3 and used read mapping to scaffolds carrying this gene to follow organisms over the 5-day eruption cycle. This approach allowed us to track 914 putatively distinct

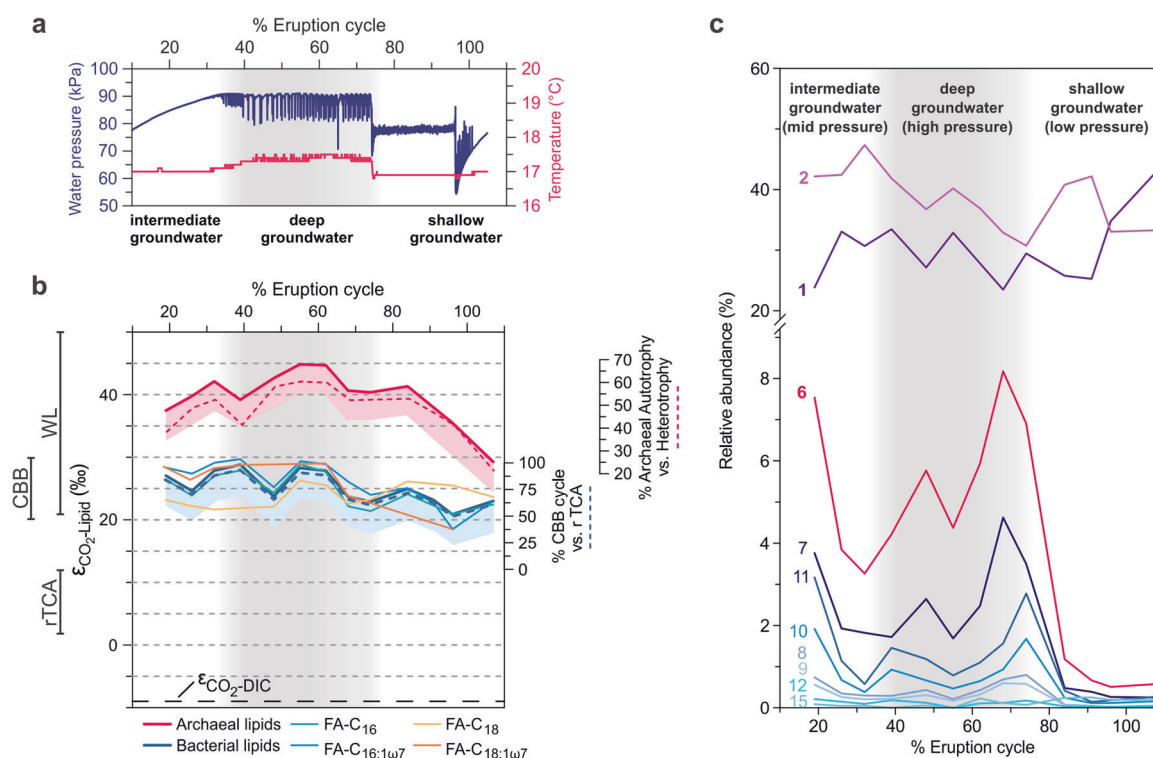
microbial species (Fig. 1), greatly exceeding the 505 previously reconstructed genomes [12].

We detected a large community shift associated with different eruption phases. According to previously published geochemical data [12], the first phase, referred to as the recovery phase, sources groundwater from an aquifer of intermediate depth, likely a Navajo Sandstone-hosted aquifer. During the second minor eruption phase, water from a deeper aquifer is sourced (likely Wingate Sandstone-hosted) and during the third major eruption phase, an increased fraction of shallow groundwater is sourced (Fig. S1). Grouping of samples into different clusters in an ordination analysis based on community composition (Fig. 1) revealed stepwise changes throughout the eruption cycle. The final sample, which was taken after the end of the major eruption phase and as the geyser transitions into the next recovery phase was size-fractionated, with cells collected sequentially on a 0.2  $\mu$ m filter and followed by a 0.1  $\mu$ m filter (sample 26, Fig. 1). The community composition on the 0.2  $\mu$ m filter plots near samples from the beginning of the first cycle in the ordination analysis, indicative of a restoration of the initial microbial community (Fig. 1).



**Fig. 1** Community structure of 27 metagenomic samples from Crystal Geyser based on percent relative abundance of scaffolds carrying rpS3 sequences (clustered at 99% amino acid similarity). Nonmetric multidimensional scaling based on the Bray–Curtis index. The connections show the trajectory of the different samples taken throughout the eruption cycle. Sample 01 was not included as it was an amplified library due to low biomass (see “Material and methods” for further details). Sample 26 was collected after the end of the major eruptions and is already part of the recovery phase (thus colored in pink). Black color indicates samples that were collected during transition between phases. Please note, that the sample was also size-fractionated into a 0.2- $\mu$ m and a 0.1- $\mu$ m filter. For details on individual rpS3 abundances please see Fig. S5 and Table S5.





**Fig. 2 Carbon isotopic ratios and relative abundance of unsaturated intact polar lipids relative to the cycle of the geyser.** **a** Water pressure and temperature over the geyser cycle showing sourcing of fluids from the conduit (mixed), the deep aquifer, and the shallow aquifer from Ref. [12]. **b** Stable carbon isotope fractionation of archaeal lipids (phytane, released from archaeol), individual bacterial fatty acids (FA, released from diacylglycerols), bacterial lipids (weighted average of FA), and dissolved inorganic carbon (DIC) relative to CO<sub>2</sub> ( $\epsilon_{\text{CO}_2\text{-Lipid}}$ ) over the geyser cycle. Lines to the left of the panel show expected ranges of  $\epsilon_{\text{CO}_2\text{-Lipid}}$  (accounting for up to 5‰ additional <sup>13</sup>C-depletion of lipids relative to biomass, indicated by shaded areas) for the Calvin-Bassham-Benson (CBB; [46–48]), the reductive tricarboxylic acid cycle (rTCA [46, 49, 50]), and the Wood-Ljungdahl pathway (WL, reductive acetyl-coenzyme A pathway; [39, 62, 63]). The blue dashed line indicates relative contribution

of carbon fixation through the CBB cycle versus the rTCA cycle for bacterial lipids (assuming maximum fractionation due to high in situ [CO<sub>2</sub>] and [DIC]). The red dashed line indicates the relative contribution of autotrophy versus heterotrophy (uptake of bacterial CBB/rTCA-fixed carbon) to archaeal lipid biomass, calculated from mass balance of  $\delta^{13}\text{C}$  values of bacterial and archaeal lipids (assuming maximum fractionation for archaeal autotrophy due to high in situ [CO<sub>2</sub>] and [DIC]). **c** Relative abundance of unsaturated diacylglycerol membrane lipids (the number indicates the sum of double bonds in both acyl chains). The distribution is dominated by mono- and diunsaturated diacylglycerols but polyunsaturated lipids (6–15 unsaturations) increase markedly in deep aquifer fluids. Grey shading indicates major eruptions, which source deep aquifer water under high pressure.

### In situ carbon fixation sustains microbial communities irrespective of aquifer depth

Previous community-wide genomic analyses suggested that carbon fixation might sustain the relatively complex aquifer microbial communities, but direct evidence was lacking [10, 33]. We measured the stable carbon isotope composition (i.e.,  $\delta^{13}\text{C}$  values) of IPL-derived bacterial FA and archaeal phytane. The values for 14 samples were plotted as a function of sampling time and compared with the  $\delta^{13}\text{C}$  values of DIC and CO<sub>2</sub> in the ecosystem (Fig. 2a, b). The  $\delta^{13}\text{C}$  values for DIC sampled from the geyser discharge over its 5-day cycle ranged from 3.6 to 8.0‰ (average = 5.0‰, std. dev. = 1.4‰) and showed no systematic variation with relative depth of source water (Fig. S6). The  $\delta^{13}\text{C}$  values for phytane range between –47.0 and –32.8‰ and for bacterial lipids (expressed as weighted average of all

FAs) from –32.7 to –22.1‰. We found very little genomic evidence for utilization of methane [35] by these communities and methane was not detected in the geyser gas emissions [12]. Thus, we do not attribute the <sup>13</sup>C-depletion of phytane to methane metabolism by methanogens/methanotrophs. Alternatively, heterotrophy could sustain microbial metabolism in the aquifers sourcing Crystal Geyser. However, the Wingate and Navajo aeolian sandstone aquifers have little associated sedimentary organic carbon [40, 41] that could serve as substrate. Similarly, dissolved organic carbon (DOC) concentrations in minor eruption phase fluids (~1 ppm, Table S3) are overall similar to global median groundwater [42], suggesting no significant admixture of exogenous DOC, for example from nearby oil reservoirs. It is possible that advection of exogenous DOC is more prevalent during major eruptions, but no DOC samples could be obtained from this phase. Still,

the  $^{13}\text{C}$ -depletion in FA and phytane suggests that the majority of biomass is not primarily derived from heterotrophic incorporation of DOC during minor eruptions: phytane ( $\delta^{13}\text{C}$  is  $-42$  to  $-47\text{‰}$ ) and the  $\text{C}_{16:0}$ ,  $\text{C}_{16:1\text{w}7}$ , and  $\text{C}_{18:1\text{w}7}$  FA are too depleted in  $^{13}\text{C}$  ( $\delta^{13}\text{C}$  is  $-29$  to  $-34\text{‰}$ ) to originate primarily from DOC ( $\delta^{13}\text{C}$  is  $-19$  to  $-24\text{‰}$ , Table S3), while only the  $\delta^{13}\text{C}$  value of  $\text{C}_{18:0}$  FA ( $-27\text{‰}$ ) is compatible with the small fractionation between substrate and FA observed in heterotrophic bacteria [43]. Importantly, the DOC in Crystal Geyser aquifers could be derived from in situ primary production and thus sustain heterotrophic bacteria.

Stable carbon isotopic compositions of lipids point to a predominantly autotrophic origin of microbial biomass. Due to the high in situ concentration of both  $\text{HCO}_3^-$  (69–84 mmol/L; [44]) and  $\text{CO}_2$  (at saturation level throughout the geyser [12, 44]), maximum fractionation by carbon-fixing microorganisms in the geyser can be assumed [45]. Changes in inorganic carbon speciation and thus fractionation are unlikely, as  $\text{HCO}_3^-$  concentrations, temperature ( $\sim 16.8$ – $17.5^\circ\text{C}$ , Fig. 2a), ionic strength (15–19 mS/cm, Fig. S1), and pH (6.4–6.5, [44]) stay in narrow ranges. In addition, growth rates of Crystal Geyser communities are likely to be low and thus carbon isotope fractionation ( $\epsilon$ ) would be expected to be maximally expressed. Based on this, and the known range of  $\epsilon$  for carbon fixed via different pathways [46–51], it is plausible that the majority of archaeal lipids were synthesized via the Wood–Ljungdahl (WL, reductive acetyl-CoA,  $\epsilon\text{DIC-lipid} > 30\text{‰}$ ) pathway from DIC, with  $\epsilon\text{DIC-lipid}$  of 38.3–53.9‰ observed in phytane derived from archaeol-based IPLs (Fig. 2b). This is in accordance with previous investigations of Crystal Geyser, which reported dominance of Altiarchaeota in the deepest aquifer [12]. Altiarchaeota fix carbon via a variant of the WL pathway with a fractionation  $\epsilon\text{DIC-lipid}$  of  $\sim 63\text{‰}$  [52] (assuming  $\epsilon\text{DIC-CO}_2$  as  $\sim 10\text{‰}$  at  $15^\circ\text{C}$  calculated after Ref. [53] and  $\delta^{13}\text{C}$  of archaeal lipids from a 99‰ enrichment of Altiarchaeota reported in Ref. [52]). The observed  $\epsilon\text{DIC-lipid}$  values for archaeal lipids in many samples are below the maximum theoretical fractionation, implying that archaea in Crystal Geyser are not exclusively autotrophic but also take up isotopically heavier organic carbon. One likely source is archaeal utilization of organic carbon fixed by bacteria via the Calvin–Benson–Bassham (CBB) and reductive tricarboxylic acid cycle (rTCA) cycles, which would be more enriched in  $^{13}\text{C}$  than carbon fixed via the WL pathway. The degree of heterotrophic uptake by archaea can be approximated using a mass balance calculation involving mixtures of carbon with (i) the maximum theoretical fractionation for autotrophic archaeal carbon fixation via the WL pathway and (ii) the observed  $\delta^{13}\text{C}$  values of bacterial lipids (accounting for up to 5‰ additional  $^{13}\text{C}$ -depletion of lipids relative to biomass). This

calculation would imply that archaea are predominantly autotrophic in deep groundwater (up to 70% of the biomass carbon fixed through WL pathway), but in the intermediate and shallow groundwater form up to 69% of their biomass by taking up bacterial organic carbon fixed through the CBB and rTCA cycles (Fig. 2b).

Bacterial lipids display the carbon isotopic fractionation expected from the CBB cycle relative to  $\text{CO}_2$  ( $\epsilon\text{CO}_2\text{-lipid}$  of 20.9–28.8‰ observed vs. 30‰ theoretical) and not that expected from fixation via the rTCA cycle ( $\epsilon\text{CO}_2\text{-lipid} < 12\text{‰}$  theoretical). Sequences encoding the CBB pathway are fairly abundant in the ecosystem throughout the recovery phase [12] and likely contributed to the bacterial lipid pool of samples collected during that period. This agrees with previous genomic findings that identified several highly active iron-oxidizing *Gallionella* species carrying this pathway [12]. Importance of *Gallionella* in Colorado Plateau aquifers is further indicated by the association of organic carbon with fossilized *Gallionella* cells in post-depositional iron concretions of the Navajo sandstone [40]. However, genomic analyses suggested that one of the most abundant organisms in the shallow aquifer (*Sulfurimonas* sp.) fixes carbon via the rTCA cycle [12]. From mass balance calculations using the observed and theoretical fractionations, we estimate that carbon fixed via the rTCA cycle contributes as little as 12% to the bacterial biomass in the deep and intermediate aquifer but up to 78% of the biomass in the shallow aquifer (Fig. 2b). Overall, the observed carbon isotopic composition of the bacterial lipids could be explained as the result of a mixture of *Sulfurimonas*-derived lipids and lipids formed via the CBB pathway.

### Degree of unsaturation of bacterial IPL changes with groundwater source depth

Using in-depth analyses of IPLs we tracked the abundance of IPL-bound bacterial unsaturated FAs across the eruption cycle. The unsaturations presumably correspond to double bonds but due to the mode of detection, we cannot strictly rule out cycloalkyl groups found in FAs of some bacteria [54], although typically not in higher numbers than one per FA. Interestingly, the relative abundance of highly unsaturated FAs correlated with the groundwater depth source (Fig. 2c). The cumulative abundances of IPLs with one or two double bond equivalents in their FA side chains were fairly consistent throughout the cycle, indicating little variation between the different groundwater sources. However, IPLs with seven or more unsaturations, i.e., at this high number presumably double bonds, were relatively abundant during the first phase, when groundwater was sourced from intermediate depths. These lipids were even more abundant during the middle phase, during which groundwater derives from the greatest depth, and almost undetectable in samples

collected in the final shallow groundwater eruption phase. One explanation for elevated abundance of polyunsaturated lipids is their derivation from eukaryotes [55]. The occurrence of tentatively identified DGCC-type (1,2-Diacylglycerol-3-O-carboxyhydroxymethylcholine) betaine lipids is unprecedented in bacteria and supports the presence of Eukaryotes in the ecosystem, although the pathway for generating these lipids and its phylogenetic distribution remains unknown [56]. In general, Eukaryotes have been found in the geyser [57], primarily in a sample of decayed wood added to the geyser conduit, and they have been detected by rpS3 analysis in the current study. However, they are not very abundant, and fluctuate heavily throughout the cycle (Fig. S7). Due to the pronounced abundance maxima during deep aquifer eruptions, the most likely explanation for the presence of polyunsaturated FAs is their origin from organisms adapted to high pressures in the deeper subsurface. Bacteria are an additional, potential source of polyunsaturated FA, as the biosynthetic capacity for these lipids is widespread in terrestrial and aquatic bacteria such as *Shewanella*, *Vibrio*, and *Geobacter* spp. [58–60]. Incorporation of double bonds in bacterial FAs is a well-known mechanism that increases membrane fluidity at high pressure and low temperature [61, 62]. Consequently, a great diversity of unsaturated FA biosynthesis gene sequences are found in the Crystal Geyser metagenomes. For instance, we detected 1959 different protein clusters (>10% dissimilarity) of 3-oxoacyl reductases, representing 11,548 protein sequences in total (Fig. S8). As temperature remained nearly constant at around 17 °C (Fig. 2), high-FA unsaturation could represent an adaptation to the high pressures faced by indigenous bacterial communities in the intermediate and deep aquifers, supporting a direct link between groundwater sources and lipid profiles.

### Predicting linkage of IPLs to uncultivated organisms

We detected 295 different IPLs in the 14 lipidomes but a strict organism-lipid relation was unresolved due to the complexity of the community. Assignment of lipids to specific organisms is further complicated by the existence of multiple potential source organisms for common lipid types and the distinct characteristics of a low-energy habitats in the subterranean aquifers. Distinct turnover times of lipids and DNA as well as lipid recycling, which may be a common strategy utilized by energy-starved archaea in the subsurface [63–65], could adversely affect correlations. While relative turnover times of DNA and lipids remain unconstrained, the predominance of chemically labile phosphoester IPLs in Crystal Geyser facilitates comparatively faster turnover of lipids compared with marine deep biosphere environments where ether-based IPLs, including glycolipids, are prevalent [66, 67]. Irrespective of whether

they represent snapshots of a dynamic system or signals accumulated over longer timescales, the systematic changes in metagenomes and lipidomes indicate distinct, stratified habitats within Crystal Geyser.

In the current study, we used a time series of 14 metagenomic and coupled lipidomic data sets to establish correlations between marker gene abundances and IPLs. Based on this analysis, we tested for evidence for the assignment of lipids to organisms. Specifically, relative abundance patterns of individual organisms were correlated with the relative abundance of the 295 IPLs (only organisms and lipids were considered if they were identified in at least seven out of fourteen samples). Lipids were also co-correlated with other lipids and primary and secondary lipid assignments were investigated via a network analysis (Fig. 3). Although the majority of IPLs were found to be unspecific, significant correlations were observed between a subset of lipids and organisms: 44 primary lipids correlate significantly with 22 different marker genes (organisms) and 63 secondary lipids (Table 1).

It is important to note that all significantly correlating ether-based isoprenoid lipids were assigned to archaea (*Ca. Huberiarchaeum crystalense*) as this provides confidence in the correlation-based approach. However, it is unclear whether correlation of the main lipid of the *Ca. H. crystalense* and one bacterial lipid is a spurious covariation or if this represents assimilation of a bacterial membrane lipid by archaea (*Huberiarchaeum* did not correlate with that bacterial lipid; Table 1). Of particular interest were the lipids of Altiarchaeota, since these had been characterized earlier [52]. These previously detected lipids, including hexose-pentose archaeol (1G-1pentose-AR; for mass spectrometric identification see Fig. S4) and dihexose extended archaeol (2G-ext-AR), were the most abundant archaeal lipids in the current study but most abundances showed little correlation with the Altiarchaeota abundances. On the one hand, this might be due to the presence of multiple different strains of *Altiarchaeum* sp. in the samples (based on rpS3 genes; Fig. S9), which can harbor different lipid profiles as shown previously [68]. On the other hand, the main archaeal IPL (2G-AR) was also present in the sample filtered through a 0.2-μm filter and collected onto a 0.1-μm filter but *Altiarchaeum* sp. DNA was not (based on rpS3 genes). This indicates the lysis of *Altiarchaeum* sp. during the filtration process, possibly due to oxygen stress, a resistance that Altiarchaeota in Crystal Geyser apparently do not possess [52]. Altiarchaeota in Crystal Geyser also have *Ca. H. crystalense* as a symbiotic partner [69], which could derive its lipids from the Altiarchaeota and was indicated to possess genes, whose products might be involved in lysis of Altiarchaeota cells [12]. In addition, longer turnover times of the chemically stable ether-bound lipids of archaea [66, 67] compared with DNA could deteriorate correlations.





**Table 1** Correlation of rpS3 gene abundances from metagenomic read mapping with relative abundance of IPL signatures across samples. Primary lipids are direct correlations, secondary lipids are those that correlated with primary lipids.

Organism classification based on rpS3 gene	Primary lipid	Secondary lipid
<i>Ca. Huberarchaeum crystalense</i>	2G-1uns-ext-AR	1G-1pentose-AR, 2G-ext-AR, C30:0-PG
<i>Ca. Forterreia multitranposorum</i>	C36:4-PC, C44:12-DGCC	
Bacteria	C44:12-DGTA	C32:2-DGTA, C30:1-DGTA, C32:3-DGCC, C32:3-DGTA, C32:4-DGTA, C34:4-DGCC, C34:5-DGTA, C34:6-DGTA, C36:5-DGCC, C36:6-DGCC, C36:6-DGTA, C36:7-DGTA, C36:8-DGTA, C37:2-DGTA, C38:7-DGCC, C40:10-DGTA, C42:11-DGTA, C42:11-PC, C36:7-DGCC
Bacteria	C63:3-DPG	
Bacteria	lyso_C18:2-DGTA	
Bacteroidetes	C36:3-PE, C36:4-PE	C30:1-PE, C34:3-PE, C34:4-PE, C30:2-PE, C32:4-PE, C32:5-PE
Citromicrobium	C32:1-2Gly-DAG, C32:2-2Gly-DAG, C34:4-2Gly-DAG, C34:5-2Gly-DAG, C34:6-2Gly-DAG, C36:6-2Gly-DAG, C36:7-2Gly-DAG, lyso_C20:4-DGCC	C38:7-2Gly-DAG, C42:11-DGCC, C30:1-2Gly-DAG, C30:2-2Gly-DAG, C32:3-2Gly-DAG, C38:9-DGCC, C61:0-DPG, C38:8-2Gly-DAG
<i>Flavobacteriaceae</i>	C63:4-DPG	C29:0-PE
<i>Flavobacteriaceae</i>	C65:4-DPG	
<i>Gallionellaceae</i>	C25:0-OL, C26:0-OL, C26:1-OL, C27:0-OL, C27:1-OL, C29:1-OL, C30:0-OL	C30:1-OL, C32:2-OL, C28:1-OL, C28:0-OL, C34:2-OL, C29:0-OL, C28:1-PE, C30:2-OL, C33:2-PE
<i>Gallionellaceae</i>	C38:6-PC	C34:1-DGCC, C34:2-PC, AEG_C36:2-2G
<i>Gallionellaceae</i>	C62:3-DPG, C64:3-DPG, C66:4-DPG	C33:3-DGTS, C68:4-DPG
<i>Hydrogenophilaceae</i>	C60:3-DPG, C64:4-DPG	
Proteobacteria	C58:1-DPG, C62:4-DPG	
Proteobacteria	lyso_C21:4-DGTS	C30:2-DGTS
Proteobacteria	lyso_C22:6-DGTS	lyso_C21:5-DGTS, lyso_C16:2-DGTS
Falkowbacteria (OD1)	lyso_C16:0-DGTS	
Kaiserbacteria (OD1)	lyso_C16:0-DGTA	lyso_C16:0-DGCC
Kaiserbacteria (OD1)	lyso_C20:4-DGTS	lyso_C16:1-DGTS, lyso_C18:3-DGTS
Parcubacteria	C30:0-DGCC	lyso_C16:3-DGTS
Parcubacteria	C47:15-DGTA	C36:5-PG, C36:4-PG
Parcubacteria	lyso_C16:2-DGCC, lyso_C18:3-DGCC, lyso_C18:4-DGCC, lyso_C20:5-DGCC	lyso_C20:5-DGTA, lyso_C16:1-DGTA, lyso_C16:1-DGCC

have the genes required to synthesize lipids from scavenged isopentenylpyrophosphate [12]. The molecular structure of 2G-1uns-ext-AR differs by only one double bond from the Altiarchaeota lipid 2G-ext-AR, so Huberarchaeota may largely derive its lipids from Altiarchaeota, which was suggested to be its host [12]. The relative abundance of 2G-1uns-ext-AR correlated significantly with 2G-ext-AR, highlighting the potential biological meaning that can be inferred from IPLs, whose abundances do not correlate with certain organisms but with certain lipids instead. Given the confident assignment of 2G-1uns-ext-AR to Huberarchaeota, we used the *p* value for that assignment as a conservative correlation *p* value for further predictions

(Bonferroni-corrected *p* value < 0.005), which are presented in Table 1.

Several bacterial groups were correlated with the occurrence of cardiolipins (diphosphatidylglycerol (DPG) lipids), which are involved in osmotic stress response, membrane ordering, and regulation of cell curvature [70–73]. Specifically, DPGs are required for maintaining cell shape in rod-shaped bacteria [71]. Consequently, DPGs found in Crystal Geyser are correlated with clades typically forming rods or elongated cell shapes, including the Flavobacteriaceae and Gallionellaceae (Table 1). These matching correlations thus further validate our statistical approach.

## Lysolipids and Candidate Phyla Radiation bacteria

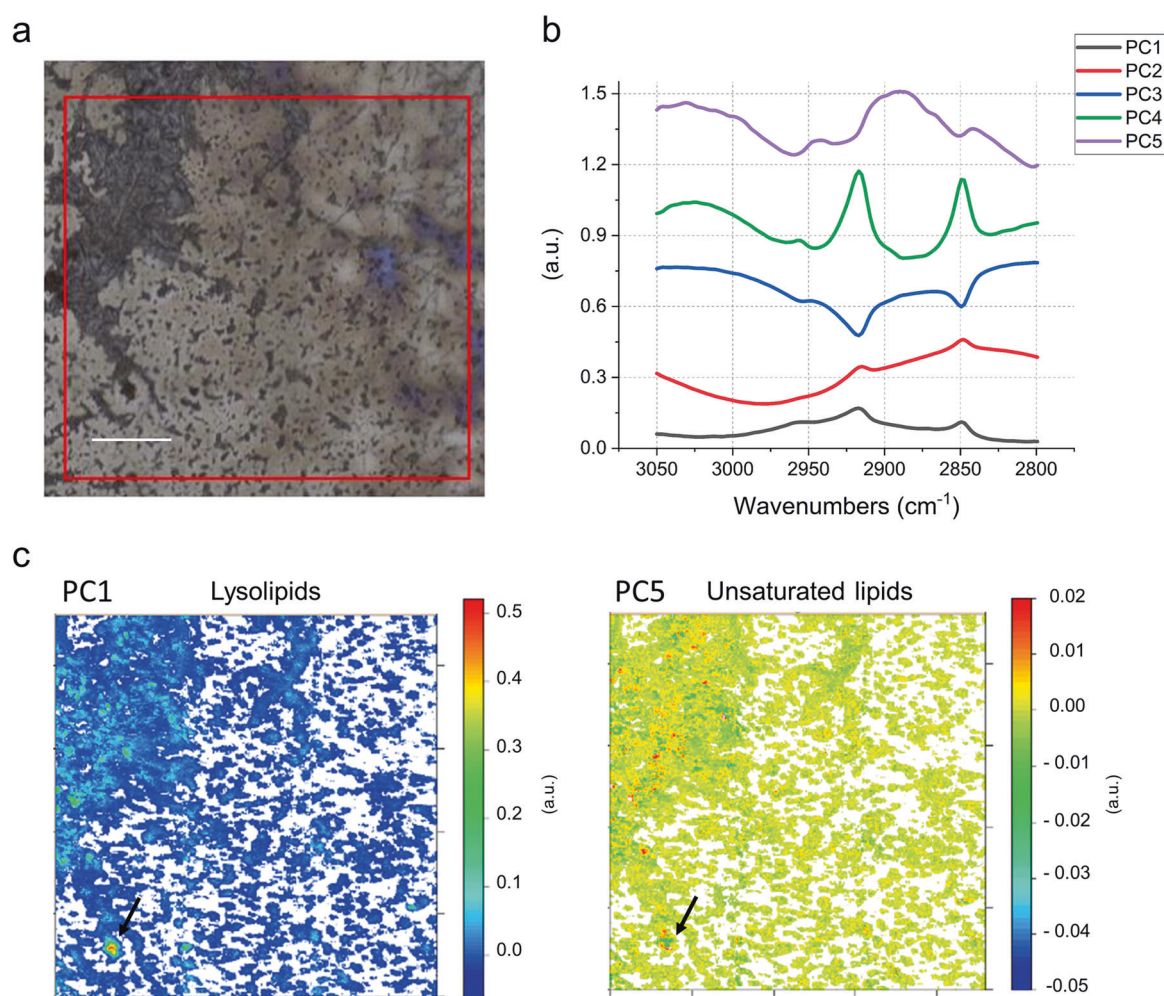
In order to investigate lipids of bacteria from the CPR [13], we analyzed the IPLs of a small cell size fraction collected on a 0.1- $\mu\text{m}$  pore-size filter after 0.2- $\mu\text{m}$  pre-filtration. Based on the corresponding metagenome, the sample contained 186 different organisms, 165 of which were classified as CPR based on rps3 sequences and one low abundant organism was classified as a member of the DPANN radiation (*Ca. H. crystalense*). Surprisingly, the most abundant organism in the sample based on metagenomics was a *Sulfurimonas*, which apparently passed through the 0.2- $\mu\text{m}$  filter (read mapping-based coverage in 0.2- $\mu\text{m}$  filter was 8.4 in the corresponding 0.1  $\mu\text{m}$  filter 1081.9). We identified 72 different IPLs in the post-0.2- $\mu\text{m}$  sample, all of which were acylglycerols. Consequently, the CPR organisms in this sample must possess FA-based lipids. This is important because the composition of lipids of CPR bacteria is unknown. Interestingly, 22 of the 72 lipids (31%) were lysolipids, all of which contained betaine headgroups (for structural characterization see Fig. S3). By contrast, these lipids constituted only 18% across the entire sample set. Cultured bacteria only contain a small fraction of lysolipids, e.g., *Sulfurimonas* has been reported to only contain a single lysolipid with ~4% abundance [74]. Further, the abundances of several CPR bacteria also correlated significantly with the abundance of specific lysolipids (Table 1).

To further investigate the lysolipid content of CPR bacteria, we selected a sample taken during the recovery phase of the geyser, when little amounts of *Sulfurimonas* are present as indicated by metagenome sequencing [12]. For this sample cells that passed through a 0.2- $\mu\text{m}$  filter were collected onto a 0.1- $\mu\text{m}$  filter for subsequent metagenomic sequencing and infrared spectromicroscopy. Metagenomic sequencing analysis of the selected sample (CG10\_big\_fil\_rev\_8\_21\_14\_0.10; [12]) showed a high abundance of CPR (rank abundance curve in Fig. S10) occupying the first seven ranks of the community. To test for the abundance of lysolipids in this CPR-rich sample, we performed FTIR analysis of the cells (Fig. 4a) and compared the results against a set of reference spectra (Fig. S11). For the first PCA in the 3050–2800  $\text{cm}^{-1}$  spectral region dominated by the aliphatic chains of the lipids, ~85% of the spectral variance is explained by the first five loading vectors (Fig. 4b). Here, the first loading vector contains 55% of the variance, with features that are similar to palmitic acid; with the asymmetric stretching of the  $\text{CH}_2$  peak centered at 2916  $\text{cm}^{-1}$  (Fig. 4). The peak corresponding to the  $\text{CH}_3$  asymmetric stretching vibration was used to evaluate the nature of the polar head. The position at 2951  $\text{cm}^{-1}$  of the PC1 is in accordance with the one of lyso-phosphatidylcholine, whereas the peak for

phosphatidylcholine is sharper and centered at 2957  $\text{cm}^{-1}$ . The corresponding heatmap of the PC1 scores (Fig. 4c) shows the presence of hotspots, a few microns in diameter. The remaining 2, 3, and 4 loading vectors, which explain 18, 10, and 2% of the variance, respectively, show different  $\text{CH}_3$  to  $\text{CH}_2$  ratios, and PC3 in particular can be assigned to free FA. In contrast, although the fifth loading vector accounts for only 1% of the variance, its spectral features can be assigned to highly branched and unsaturated lipids similar to those of archaea (Fig. 4b, c; Refs. [14, 75]; see Supplementary material for additional results). This agrees with the presence of DPANN archaea as the second most prominent group of organisms in this sample based on metagenomic profiling (Fig. S10). The combination of the detailed analysis of the IPLs and infrared imaging of two independently sampled small cell fractions suggest that a substantial fraction of some CPR cell membranes consists of lysolipids.

Genome-resolved metagenomics generated 206 new genomes from the entire sample set. Together with 1215 previous genomes [12, 35], our data set included 675 genomes of CPR bacteria that were used to comprehensively investigate their potential for lipid biosynthesis (accessible through [https://ggkbase.berkeley.edu/genome\\_summaries/1491-Bacterial\\_membrane\\_lipids\\_AJP](https://ggkbase.berkeley.edu/genome_summaries/1491-Bacterial_membrane_lipids_AJP)). We found that the CPR genomes do not encode for any known, complete bacterial lipid biosynthesis pathway, yet CPR bacteria are known to have a cytoplasmic membrane based on cryogenic-transmission electron microscopy studies [14]. Interestingly, some members of the Neelsonbacteria phylum (Parcubacteria superphylum) have near-complete pathways for FAs and phospholipid synthesis. They possess some homologs of the FA synthase type II (FAS-II), the main FA biosynthesis pathway in most bacteria. However, they lack the FAS-related acyl carrier protein (ACP) processing machinery (ACP synthase and malonyl-CoA:ACP transacylase). ACP is a peptide cofactor that functions as a shuttle that covalently binds all FA intermediates. Although they lack key genes for FA synthesis, we cannot rule out this group could potentially synthesize FAs by an ACP-independent pathway, as suggested for some archaea [76]. We also searched these genomes for genes coding for glycerol-3-phosphate (G3P) dehydrogenase, an enzyme responsible for the stereochemistry of the glycerol units of their membrane lipids, and acyl-ACP transferases responsible for the formation of ester bonds between FAs and G3P backbone in phospholipid synthesis. There are two families of acyltransferases responsible for the acylation of the C1-position of the G3P. The PlsB acyltransferase primarily uses ACP end products of FA biosynthesis (acyl-ACP) as acyl donors. The second family involves the PlsY acyltransferase and is more widely distributed in Bacteria. PlsY uses as



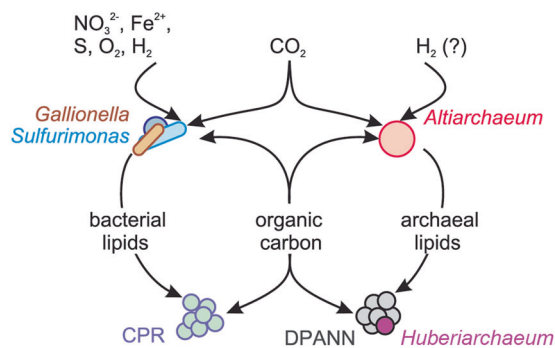


**Fig. 4** FTIR analysis of a small cell size fraction (post-0.2- $\mu\text{m}$  filter collected onto a 0.1- $\mu\text{m}$  filter). **a** Field of view in FTIR,  $1 \times 1$  mm (red square). **b** First five PCA loadings accounting for  $\sim 90\%$  of the variance. They describe the directions of maximum variability of the analyzed system. The figure presets the first five vectors, that spectroscopically can be assigned, by similarity of shape and band position, to different types of lipids. **c** False color maps representing PCA scores PC1 and PC5, respectively. These maps show how the different lipids represented by the eigenvectors in (b), are distributed in the sample. The comparison of the spectral features of the loadings and the reference spectra in Fig. S11 allow assignment of PC1 to lysolipids

and PC5 to unsaturated/branched lipids. The arrows point to a hotspot of cells indicating a particularly high distribution of lysolipids (PC1), surrounded by several smaller hotspots of unsaturated/branched lipids (PC5). Given the micrometric lateral resolution of the image (each pixel is  $2.6 \mu\text{m}$ ) it is possible to hypothesize that there is a small group of cells in the hotspot area, which is characterized by distinct membrane lipid composition. This can also be observed in other spots throughout the measured biomass. Loadings of the PCA over the whole  $900\text{--}3700 \text{ cm}^{-1}$  spectral range are provided in Fig. S12. Scale bar  $200 \mu\text{m}$ .

donor acyl-phosphate produced from acyl-ACP by PlsX (an acyl-ACP:PO<sub>4</sub> transacylase enzyme). The acylation in the C2-position of the G3P is carried out by the 1-acylglycerol-3-phosphate O-acyltransferase (PlsC). Screening the Nealonbacteria genomes, we did not detect any homologs of the first family of acyltransferase, PlsB. However, we identified PlsY and PlsC, but not PlsX. Absence of PlsX raises the question of the enzyme or mechanism for production of acyl-phosphate needed to activate PlsY. Overall, mechanisms or enzymes that produce and/or require ACP were not identified in CPR genomes in this study. Even though this

finding opens the possibility for the presence of ACP-independent pathways for FA and/or lipid synthesis in these CPR bacteria, we cannot conclude with confidence that few of these organisms can synthesize lipids *de novo*. Thus, we suggest that most CPR bacteria derive their membrane lipids, including lysolipids, from coexisting bacteria. Given the small cell size of CPR, lysolipids may be preferred due to their role in reducing membrane curvature stress (e.g., Ref. [77]). As lysolipids can form during lipid breakdown (e.g., mediated by phospholipase A [78]) and can be taken up by other bacteria [79], their utilization by CPR may



**Fig. 5 Model for the acquisition and redistribution of carbon and lipids in the deep subsurface ecosystems of the Colorado Plateau (USA) accessible through Crystal Geyser.** Organic carbon and lipids are produced by *Gallionella*, *Sulfurimonas*, *Altiarchaeum* spp. or other autotrophs, redistributed through the ecosystem and acquired by other community members including CPR bacteria and DPANN archaea.

indicate uptake from degraded bacterial biomass or direct derivation from host cells.

### Model of lipid transfer in the community and conclusions

Our approach combined detailed metagenomics with whole community lipidomics and infrared spectroscopy and was informed by isotopic measurements that were constrained by detailed understanding of the geological context. The objective was to probe the carbon cycle within the subsurface microbial ecosystem, particularly the source of fixed organic carbon, but also to investigate evidence for its redistribution into other organisms, especially putative symbionts. Although sample limitation resulted in a lower resolution of isotopic analyses compared with metagenomics, carbon isotope systematics of archaeal and bacterial lipids confidently support the metagenomic predictions that microbial biomass is mostly of autotrophic origin in all aquifers sampled. Particularly, our results provide evidence that predicted autotrophs were fixing  $\text{CO}_2$  in situ, using the WL (*Altiarchaeum*), rTCA (*Sulfurimonas*), and CBB cycles (*Gallionella*).

Using lipidomics and infrared spectroscopy on size-fractionated cells, we demonstrate that CPR bacteria with small cell size possess FA-based IPLs, although the corresponding genomes do not encode for a known pathway to synthesize them. Similarly, Huberarchaeota, potential symbionts of *Altiarchaeota*, were predicted to possess altered archaeal lipids related to those of their putative hosts. Our results support the notion that organisms of the CPR and DPANN radiation do not only scavenge (or symbiotically receive) molecular building blocks or even intact lipids from other bacteria and archaea but also use the corresponding lipids and introduce modifications (Fig. 5).

**Acknowledgements** We are grateful to Susan Spaulding, Christopher T Brown, Karthik Anantharaman, Ken Yu Khaw, and Ilona Ruhl for assistance with sampling. We thank Julius Lipp for providing archaeal lipid standards and supporting lipid analyses. This study was funded by the Sloan Foundation (“Deep Life,” grant no. G-2016-20166041). AJP was also supported by the Deutsche Forschungsgemeinschaft (DFG PR 1603/1-1) and acknowledges funding by the Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen (“Nachwuchsgruppe, AJP”). Lipid extractions and analyses at the University of Bremen were supported by the Deutsche Forschungsgemeinschaft through the Gottfried Wilhelm Leibniz Program (award to KUH; Hi 616-14-1) and a project grant to JFB, AJP, and KUH by the “Deep Life Community” of the “Deep Carbon Observatory”, which is supported by the Alfred P. Sloan Foundation. MCR acknowledges funding by the Canadian National Science and Engineering Research Council (“Discovery Grant”, 06509-2016). The work conducted by the US Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, and by DOE’s Berkeley Synchrotron Infrared Structural BioImaging (BSISB) Resource at the Advanced Light Source are supported under Contract No. DE-AC02-05CH11231. Open access funding provided by Projekt DEAL.

### Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

### References

- Suzuki S, Ishii S, Hoshino T, Rietze A, Tenney A, Morrill PL, et al. Unusual metabolic diversity of hyperalkaliphilic microbial communities associated with subterranean serpentinization at The Cedars. *ISME J.* 2017;11:2584–98.
- Hernsdorf AW, Amano Y, Miyakawa K, Ise K, Suzuki Y, Anantharaman K, et al. Potential for microbial  $\text{H}_2$  and metal transformations associated with novel bacteria and archaea in deep terrestrial subsurface sediments. *ISME J.* 2017;11:1915–29.
- Hu P, Tom L, Singh A, Thomas BC, Baker BJ, Piceno YM, et al. Genome-resolved metagenomic analysis reveals roles for candidate phyla and other microbial community members in biogeochemical transformations in oil reservoirs. *MBio.* 2016;7:e01669–15.
- Orsi WD. Ecology and evolution of seafloor and subseafloor microbial communities. *Nat Rev Microbiol.* 2018;16:671–83.
- Magnabosco C, Lin L-H, Dong H, Bomberg M, Ghiorse W, Stan-Lotter H, et al. The biomass and biodiversity of the continental subsurface. *Nat Geosci.* 2018;11:707.



6. Magnabosco C, Ryan K, Lau MCY, Kuloyo O, Sherwood Lollar B, Kieft TL, et al. A metagenomic window into carbon metabolism at 3 km depth in Precambrian continental crust. *ISME J*. 2016;10:730–41.
7. Chivian D, Brodie EL, Alm EJ, Culley DE, Dehal PS, DeSantis TZ, et al. Environmental genomics reveals a single-species ecosystem deep within Earth. *Science*. 2008;322:275–8.
8. Lau MCY, Kieft TL, Kuloyo O, Linage-Alvarez B, van Heerden E, Lindsay MR, et al. An oligotrophic deep-subsurface community dependent on syntrophy is dominated by sulfur-driven autotrophic denitrifiers. *Proc Natl Acad Sci USA*. 2016;113: E7927–36.
9. Momper L, Jungbluth SP, Lee MD, Amend JP. Energy and carbon metabolisms in a deep terrestrial subsurface fluid microbial community. *ISME J*. 2017;11:2319–33.
10. Osburn MR, LaRowe DE, Momper LM, Amend JP. Chemolithotrophy in the continental deep subsurface: Sanford Underground Research Facility (SURF), USA. *Front Microbiol*. 2014;5:610.
11. Emerson JB, Thomas BC, Alvarez W, Banfield JF. Metagenomic analysis of a high carbon dioxide subsurface microbial community populated by chemolithoautotrophs and bacteria and archaea from candidate phyla: high CO<sub>2</sub> subsurface metagenomics. *Environ Microbiol*. 2016;18:1686–703.
12. Probst AJ, Ladd B, Jarett JK, Geller-McGrath DE, Sieber CMK, Emerson JB, et al. Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nat Microbiol*. 2018;3:328–36.
13. Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, et al. Unusual biology across a group comprising more than 15% of domain bacteria. *Nature*. 2015;523:208–11.
14. Luef B, Frischkorn KR, Wrighton KC, Holman HN, Birarda G, Thomas BC, et al. Diverse uncultivated ultra-small bacterial cells in groundwater. *Nat Commun*. 2015;6:1–8.
15. Castelle CJ, Banfield JF. Major new microbial groups expand diversity and alter our understanding of the tree of life. *Cell*. 2018;172:1181–97.
16. He X, McLean JS, Edlund A, Yooseph S, Hall AP, Liu S-Y, et al. Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. *Proc Natl Acad Sci USA*. 2015;112: 244–9.
17. Waters E, Hohn MJ, Ahel I, Graham DE, Adams MD, Barnstead M, et al. The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc Natl Acad Sci USA*. 2003;100:12984–8.
18. Anbutu H, Moriyama M, Nikoh N, Hosokawa T, Futahashi R, Tanahashi M, et al. Small genome symbiont underlies cuticle hardness in beetles. *Proc Natl Acad Sci USA*. 2017;114:E8382–91.
19. Wurch L, Giannone RJ, Belisle BS, Swift C, Utturkar S, Hettich RL, et al. Genomics-informed isolation and characterization of a symbiotic Nanoarchaeota system from a terrestrial geothermal environment. *Nat Commun*. 2016;7:12115.
20. Jahn U, Summons R, Sturt H, Grosjean E, Huber H. Composition of the lipids of *Nanoarchaeum equitans* and their origin from its host *Ignicoccus* sp. strain KIN4/I. *Arch Microbiol*. 2004; 182:404–13.
21. Boetius A, Ravensschlag K, Schubert CJ, Rickert D, Widdel F, Gieseke A, et al. A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature*. 2000; 407:623–6.
22. Hinrichs K-U, Summons RE, Orphan V, Sylva SP, Hayes JM. Molecular and isotopic analysis of anaerobic methane-oxidizing communities in marine sediments. *Org Geochem*. 2000; 31:1685–701.
23. Raghoebarsing AA, Pol A, van de Pas-Schoonen KT, Smolders AJP, Ettwig KF, Rijpstra WIC, et al. A microbial consortium couples anaerobic methane oxidation to denitrification. *Nature*. 2006;440:918–21.
24. Schubotz F, Meyer-Dombard DR, Bradley AS, Fredricks HF, Hinrichs K-U, Shock EL, et al. Spatial and temporal variability of biomarkers and microbial diversity reveal metabolic and community flexibility in Streamer Biofilm Communities in the Lower Geyser Basin, Yellowstone National Park. *Geobiology*. 2013;11:549–69.
25. Schubotz F, Lipp JS, Elvert M, Hinrichs K-U. Stable carbon isotopic compositions of intact polar lipids reveal complex carbon flow patterns among hydrocarbon degrading microbial communities at the Chapopote asphalt volcano. *Geochim Cosmochim Acta*. 2011;75:4399–415.
26. Green CT, Scow KM. Analysis of phospholipid fatty acids (PLFA) to characterize microbial communities in aquifers. *Hydrogeol J*. 2000;8:126–41.
27. Sturt HF, Summons RE, Smith K, Elvert M, Hinrichs K-U. Intact polar membrane lipids in prokaryotes and sediments deciphered by high-performance liquid chromatography/electrospray ionization multistage mass spectrometry—new biomarkers for biogeochemistry and microbial ecology. *Rapid Commun Mass Spectrom*. 2004;18:617–28.
28. Wörmer L, Lipp JS, Schröder JM, Hinrichs K-U. Application of two new LC-ESI-MS methods for improved detection of intact polar lipids (IPLs) in environmental samples. *Org Geochem*. 2013;59:10–21.
29. Peng Y, Leung HC, Yiu S-M, Chin FY. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*. 2012;28:1420–8.
30. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinforma*. 2010;11:119.
31. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
32. Sharon I, Kertesz M, Hug LA, Pushkarev D, Blauwkamp TA, Castelle CJ, et al. Accurate, multi-kb reads resolve complex populations and detect rare microorganisms. *Genome Res*. 2015;25:534–43.
33. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, et al. A new view of the tree of life. *Nat Microbiol*. 2016;1:16048.
34. Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, et al. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun*. 2016;7:13219.
35. Probst AJ, Castelle CJ, Singh A, Brown CT, Anantharaman K, Sharon I, et al. Genomic resolution of a cold subsurface aquifer community provides metabolic insights for novel microbes adapted to high CO<sub>2</sub> concentrations. *Environ Microbiol*. 2017;19:459–74.
36. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004; 32:1792–7.
37. Wrighton KC, Thomas BC, Sharon I, Miller CS, Castelle CJ, VerBerkmoes NC, et al. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science*. 2012;337:1661–5.
38. Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics*. 2007;23:1282–8.
39. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 1999;27:29–34.
40. Weber KA, Spanbauer TL, Wacey D, Kilburn MR, Loope DB, Kettler RM. Biosignatures link microorganisms to iron mineralization in a paleoaquifer. *Geology*. 2012;40:747–50.

41. Beittler B, Parry WT, Chan MA. Fingerprints of fluid flow: chemical diagenetic history of the Jurassic Navajo Sandstone, Southern Utah, USA. *J Sediment Res.* 2005;75:547–61.
42. McDonough L, Santos I, Andersen M, O'Carroll D, Rutledge H, Meredith K, et al. Changes in global groundwater organic carbon driven by climate change and urbanization. *EarthArXiv.* 2018. <https://doi.org/10.31223/osf.io/vmaku>.
43. Blair N, Leu A, Muñoz E, Olsen J, Kwong E, Marais DD. Carbon isotopic fractionation in heterotrophic microbial metabolism. *Appl Environ Microbiol.* 1985;50:996–1001.
44. Han WS, Lu M, McPherson BJ, Keating EH, Moore J, Park E, et al. Characteristics of CO<sub>2</sub>-driven cold-water geyser, Crystal Geyser in Utah: experimental observation and mechanism analyses. *Geofluids.* 2013;13:283–97.
45. Fogel ML, Cifuentes LA. Isotope fractionation during primary production. In: Engel MH, Macko SA, (eds). *Organic geochemistry: principles and applications.* Boston, MA: Springer; 1993. p. 73–98.
46. Berg IA, Kockelkorn D, Ramos-Vera WH, Say RF, Zarzycki J, Hügler M, et al. Autotrophic carbon fixation in archaea. *Nat Rev Microbiol.* 2010;8:447–60.
47. Quandt L, Gottschalk G, Ziegler H, Stichler W. Isotope discrimination by photosynthetic bacteria. *FEMS Microbiol Lett.* 1977;1:125–8.
48. McNevin DB, Badger MR, Whitney SM, Von Caemmerer S, Tcherkez GGB, Farquhar GD. Differences in carbon isotope discrimination of three variants of D-ribulose-1,5-bisphosphate carboxylase/oxygenase reflect differences in their catalytic mechanisms. *J Biol Chem.* 2007;282:36068–76.
49. Preuß A, Schauder R, Fuchs G, Stichler W. Carbon isotope fractionation by autotrophic bacteria with three different CO<sub>2</sub> fixation pathways. *Z Naturforsch.* 1989;44:397–402.
50. Sirevåg R, Buchanan BB, Berry JA, Troughton JH. Mechanisms of CO<sub>2</sub> fixation in bacterial photosynthesis studied by the carbon isotope fractionation technique. *Arch Microbiol.* 1977;112:35–38.
51. Fuchs G. Alternative pathways of autotrophic CO<sub>2</sub> fixation. In: Schlegel HG, Bowien B, (eds). *Biology of autotrophic bacteria.* Berlin: Springer; 1989. p. 365–82.
52. Probst AJ, Weinmaier T, Raymann K, Perras A, Emerson JB, Rattei T, et al. Biology of a widespread uncultivated archaeon that contributes to carbon fixation in the subsurface. *Nat Commun.* 2014;5:5497.
53. Mook WG, Bommerson JC, Staverman WH. Carbon isotope fractionation between dissolved bicarbonate and gaseous carbon dioxide. *Earth Planet Sci Lett.* 1974;22:169–76.
54. Buist PH. Exotic biomodification of fatty acids. *Nat Prod Rep.* 2007;24:1110–27.
55. Kato M, Sakai M, Adachi K, Ikemoto H, Sano H. Distribution of betaine lipids in marine algae. *Phytochemistry.* 1996;42:1341–5.
56. Kato M, Kobayashi Y, Torii A, Yamada M. Betaine lipids in marine algae. In: Murata N, Yamada M, Nishida I, Okuyama H, Sekiya J, Hajime W, (eds). *Advanced research on plant lipids.* Dordrecht: Springer; 2003. p. 19–22.
57. West PT, Probst AJ, Grigoriev IV, Thomas BC, Banfield JF. Genome-reconstruction for eukaryotes from complex natural microbial communities. *Genome Res.* 2018;28:569–80.
58. Metz JG, Roessler P, Facciotti D, Levering C, Ditttrich F, Lassner M, et al. Production of polyunsaturated fatty acids by polyketide synthases in both prokaryotes and eukaryotes. *Science.* 2001;293:290–3.
59. Hamamoto T, Takata N, Kudo T, Horikoshi K. Characteristic presence of polyunsaturated fatty acids in marine psychrophilic vibrios. *FEMS Microbiol Lett.* 1995;129:51–56.
60. Yoshida K, Hashimoto M, Hori R, Adachi T, Okuyama H, Ori-kasa Y, et al. Bacterial long-chain polyunsaturated fatty acids: their biosynthetic genes, functions, and practical use. *Mar Drugs.* 2016;14:94.
61. Russell NJ. Psychrophilic bacteria—molecular adaptations of membrane lipids. *Comp Biochem Physiol Part A: Physiol.* 1997;118:489–93.
62. DeLong E, Yayanos A. Adaptation of the membrane lipids of a deep-sea bacterium to changes in hydrostatic pressure. *Science.* 1985;228:1101–3.
63. Liu X, Lipp JS, Hinrichs K-U. Distribution of intact and core GDGTs in marine sediments. *Org Geochem.* 2011;42:368–75.
64. Lipsewiers YA, Hopmans EC, Sinninghe Damsté JS, Villanueva L. Potential recycling of thaumarchaeotal lipids by DPANN Archaea in seasonally hypoxic surface marine sediments. *Org Geochem.* 2018;119:101–9.
65. Takano Y, Chikaraishi Y, Ogawa NO, Nomaki H, Morono Y, Inagaki F, et al. Sedimentary membrane lipids recycled by deep-sea benthic archaea. *Nat Geosci.* 2010;3:858–61.
66. Logemann J, Graue J, Köster J, Engelen B, Rullkötter J, Cypionka H. A laboratory experiment of intact polar lipid degradation in sandy sediments. *Biogeosciences.* 2011;8:2547–60.
67. Xie S, Lipp JS, Wegener G, Ferdelman TG, Hinrichs K-U. Turnover of microbial lipids in the deep biosphere and growth of benthic archaeal populations. *Proc Natl Acad Sci USA.* 2013;110:6010–4.
68. Probst AJ, Birarda G, Holman H-YN, DeSantis TZ, Wanner G, Andersen GL, et al. Coupling genetic and chemical microbiome profiling reveals heterogeneity of Archaeome and Bacteriome in subsurface biofilms that are dominated by the same Archaeal species. *PLoS ONE.* 2014;9:e99801.
69. Schwank K, Bornemann TLV, Dombrowski N, Spang A, Banfield JF, Probst AJ. An archaeal symbiont-host association from the deep terrestrial subsurface. *ISME J.* 2019;13:2135–9.
70. Romantsov T, Guan Z, Wood JM. Cardiolipin and the osmotic stress responses of bacteria. *Biochim Biophys Acta—Biomembr.* 2009;1788:2092–100.
71. Khoury ME, Swain J, Sautrey G, Zimmermann L, Smissen PVD, Décout J-L, et al. Targeting bacterial cardiolipin enriched microdomains: an antimicrobial strategy used by amphiphilic aminoglycoside antibiotics. *Sci Rep.* 2017;7:1–12.
72. Renner LD, Weibel DB. Cardiolipin microdomains localize to negatively curved regions of *Escherichia coli* membranes. *PNAS.* 2011;108:6264–9.
73. Lin T-Y, Gross WS, Auer GK, Weibel DB. Cardiolipin alters *Rhodobacter sphaeroides* cell shape by affecting peptidoglycan precursor biosynthesis. *mBio.* 2019;10:e02401–18.
74. Götz F, Longnecker K, Soule MCK, Becker KW, McNichol J, Kujawinski EB, et al. Targeted metabolomics reveals proline as a major osmolyte in the chemolithoautotroph *Sulfurimonas denitrificans*. *MicrobiologyOpen.* 2018;7:e00586.
75. Probst AJ, Holman H-YN, DeSantis TZ, Andersen GL, Birarda G, Bechtel HA, et al. Tackling the minority: sulfate-reducing bacteria in an archaea-dominated subsurface biofilm. *ISME J.* 2013;7:635.
76. Lombard J, López-García P, Moreira D. An ACP-independent fatty acid synthesis pathway in Archaea: implications for the origin of phospholipids. *Mol Biol Evol.* 2012;29:3261–5.
77. Fuller N, Rand RP. The influence of lysolipids on the spontaneous curvature and bending elasticity of phospholipid membranes. *Biophysical J.* 2001;81:243–54.
78. Sahonero-Canavesi DX, López-Lara IM, Geiger O. Aerobic utilization of hydrocarbons, oils and lipids. In: Rojo F, (ed.). *Handbook of hydrocarbon and lipid microbiology.* Cham: Springer International Publishing; 2018. p. 1–24.
79. Hsu L, Jackowski S, Rock CO. Uptake and acylation of 2-acyl-lysophospholipids by *Escherichia coli*. *J Bacteriol.* 1989;171:1203–5.